


Immature acute leukaemias: lessons from the haematopoietic roadmap

Thomas Lefeuvre^{1,2}, Luke Jones^{1,2}, Amélie Trinquand^{2,3}, Antoine Pinton^{4,5}, Elizabeth Macintyre^{4,5}, Elisa Laurenti^{6,7} and Jonathan Bond^{1,2,3} 

1 Systems Biology Ireland, School of Medicine, University College Dublin, Dublin, Ireland

2 National Children's Research Centre, Dublin, Ireland

3 Children's Health Ireland at Crumlin, Dublin, Ireland

4 Laboratory of Onco-Haematology, Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Necker Enfants-Malades, Université de Paris, Paris, France

5 Institut Necker-Enfants Malades (INEM), Institut national de la santé et de la recherche médicale (Inserm) U1151, Paris, France

6 Department of Haematology, University of Cambridge, Cambridge, UK

7 Wellcome and MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

Keywords

acute leukaemia; acute myeloid leukaemia; early thymic precursor-ALL; haematopoiesis; ICGS; mixed-phenotype acute leukaemia

Correspondence

J. Bond, Systems Biology Ireland, University College Dublin, Belfield, Dublin 4, Ireland
Tel: +353 1 716 6338
E-mail: jonathan.bond@ucd.ie

It is essential to relate the biology of acute leukaemia to normal blood cell development. In this review, we discuss how modern models of haematopoiesis might inform approaches to diagnosis and management of immature leukaemias, with a specific focus on T-lymphoid and myeloid cases. In particular, we consider whether next-generation analytical tools could provide new perspectives that could improve our understanding of immature blood cancer biology.

(Received 2 February 2021, revised 30 April 2021, accepted 20 May 2021)

doi:10.1111/febs.16030

Abbreviations

ABL1, ABL proto-oncogene 1; AL, acute leukaemia; ALAL, acute leukaemia of ambiguous lineage; ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; AMTL, acute myeloid/T-lymphoblastic leukaemia; ASXL1, ASXL transcriptional regulator 1; ATAC-seq, assay for transposase-accessible chromatin using sequencing; AUL, acute undifferentiated leukaemia; BAL, biphenotypic acute leukaemia; BCR, breakpoint cluster region gene; CITE-seq, cellular indexing of transcriptomes and epitopes by sequencing; CMP, common myeloid progenitor; *DNMT3A*, DNA methyltransferase 3 Alpha; *EED*, embryonic ectoderm development; EGIL, European Group for Immunological Classification of Leukaemias; ETP, early thymic progenitor; ETS, erythroblast transformation-specific domain; *ETV6*, ETS variant transcription factor 6; *EZH2*, enhancer of zeste 2; *FLT3*, Fms-related receptor tyrosine kinase 3; *GATA3*, GATA binding protein 3; GMP, granulocyte-monocyte progenitor; GSEA, gene set enrichment analysis; IAL, interface acute leukaemia; ICGS, iterative clustering and guide gene selection; *IL7R*, interleukin 7 receptor; JAK, Janus kinase; *KMT2A*, lysine methyltransferase 2A; LMO2, LIM domain only 2; LT-HSC, long-term haematopoietic stem cell; MARK, microtubule-associated protein (MAP)-microtubule affinity regulating kinase; *MEF2C*, myocyte enhancer factor 2C; MEP, megakaryocyte-erythrocyte progenitor; MFC, multiparameter flow cytometry; MLP, multilymphoid progenitor; MPAL, mixed-phenotype acute leukaemia; MPO, myeloperoxidase; MPP, multipotent progenitor; NOS, not otherwise specified; PCA, Principal Component Analysis; PHF6, PHD finger protein 6; PNT, N-terminal pointed; PRC2, polycomb repressive complex 2; PTEN, phosphatase and tensin homolog; SIK, salt-inducible kinases; SNE, stochastic neighbour embedding; STAT, signal transducer and activator of transcription; ST-HSC, short-term haematopoietic stem cell; TAD, transactivation domain; TAL1, TAL BHLH transcription factor 1, erythroid differentiation factor; T-ALL, T-cell acute lymphoblastic leukaemia; TdT, terminal deoxynucleotidyl transferase; TET2, Tet methylcytosine dioxygenase 2; TKD, tyrosine kinase domain; TR, T-receptor; WHO, World Health Organisation.

Acute leukaemia: a haematopoietic perspective

As acute leukaemia (AL) is caused by malignant proliferation of blood cells arrested at an immature stage of development, it is essential to relate leukaemia biology to normal haematopoietic processes. As well as providing a theoretical framework to understand oncogenesis, comparison with normal differentiation can provide rational avenues for tackling the pathological differentiation block.

AL classification has traditionally been based on historical haematopoiesis models that included an early separation into myeloid and lymphoid lineages. This means that ALs are broadly categorised as either acute myeloid (AML) or acute lymphoblastic leukaemia (ALL), with the latter being further subdivided by B- and T-lymphoid origin. In rare cases, ALs cannot be allocated to either lineage, so are classed as acute leukaemias of ambiguous lineage (ALAL). Modern haematopoietic models (Fig. 1) no longer include this early myeloid/lymphoid dichotomy, due to strong

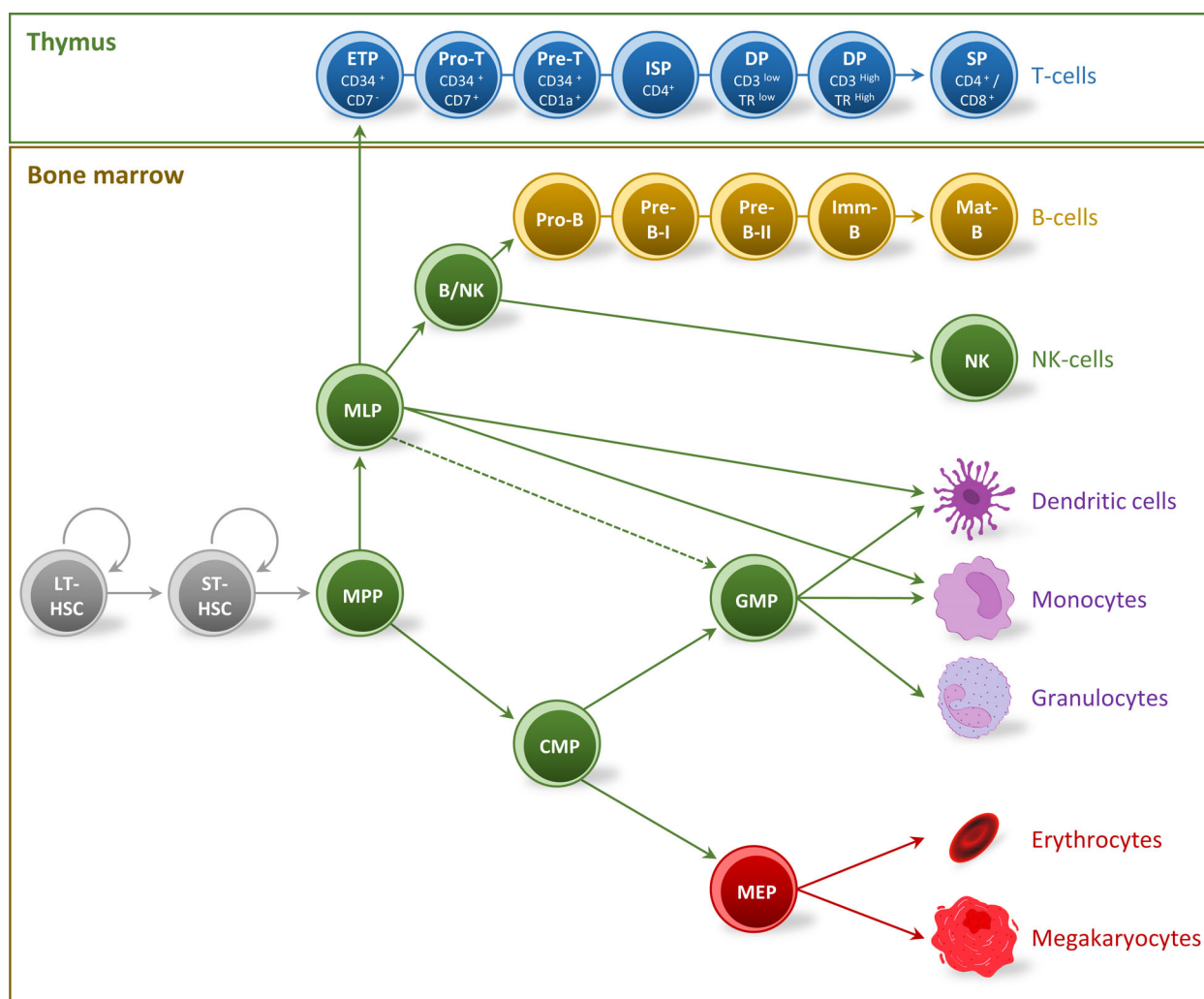


Fig. 1. Schematic representation of normal haematopoiesis. Major developmental stages for human myeloid and T-lymphoid development within the bone marrow and thymus are shown. Steady-state blood formation is largely driven by ST-HSC/MPPs [130,131] whereas LT-HSCs contribute to differentiation upon transplantation or other severe stress [132]. CMP, common myeloid progenitor; ETP, early thymic progenitor; GMP, granulocyte-monocyte progenitor; Imm. B, immature B cell; LT-HSC, long-term haematopoietic stem cell; Mat. B, mature B cell; MEP, megakaryocyte-erythrocyte progenitor; MLP, multilymphoid progenitor; MPP, multipotent progenitor; Pre-B, Pre-B cell; Pro-B, Pro-B cell; ST-HSC, short-term haematopoietic stem cell.

evidence that progenitors retain the potential to differentiate into several lineages much later than was initially realised. For example, various lymphoid cell types have been shown to have myeloid potential at even relatively advanced stages of differentiation in the bone marrow [1–6], and even the thymus [7,8].

Given that the haematopoietic roadmap has evolved, do we need to update the leukaemia cartography too? In this review, we will first discuss the limitations of current diagnostic practices in the light of contemporary knowledge of blood cell development. We will next examine how mutational genotype could earmark AL subsets that develop from similar immature progenitors. Finally, we will consider whether new tools to analyse normal haematopoiesis could provide different perspectives on leukaemia ontogeny.

Current practical challenges

Although classification now increasingly incorporates analysis of specific molecular alterations in leukaemic blasts, AL categorisation in clinical practice is still effectively determined by multiparameter flow cytometry (MFC) assessment of immunophenotypic resemblance to normal haematopoietic precursors [9]. Almost all ALs are classed as either myeloid, B-lymphoid or T-lymphoid, based on analysis of expression of cytoplasmic or surface antigens that are associated with each lineage.

Two main sets of diagnostic reference criteria have been described. First, the European Group for Immunological Classification of Leukaemias (EGIL) proposed a scoring system based on a broad panel of markers to define myeloid, B- or T-lymphoid lineage (Table 1) [10]. This system is weighted by the lineage fidelity of each antigen, with fewer points attributed to

markers that are known to be commonly aberrantly expressed (e.g. CD7). MFC positivity is defined at a single threshold of blasts with positive staining in comparison with an isotype control ($\geq 20\%$ for surface antigens and $\geq 10\%$ for cytoplasmic markers), meaning that the intensity of antigen expression in ALs can vary markedly. In contrast to the relatively broad range of EGIL-defined antigens, World Health Organization (WHO) guidelines emphasise a small number of key lineage-defining markers, in particular MPO (myeloid), CD19 (B-lymphoid) and CD3 (T-lymphoid), with no recommended positivity threshold (Table 2) [9,11,12].

These classification systems provide a theoretically simple and practical workflow to diagnose most ALs with relative certainty. However, it might be expected that leukaemias arising from transformation of nonlineage-restricted haematopoietic precursors (Fig. 1) could have patterns of antigen expression that are not so easy to pigeonhole. A small but not insignificant minority ($\leq 5\%$) of ALs cannot be clearly categorised as either myeloid or lymphoid leukaemia in practice. These cases pose a clinical conundrum, as therapies for AML and ALL are very different. There is little consensus on how best to treat these cases, and patient outlook is typically poor [13–15].

The EGIL have defined biphenotypic acute leukaemia (BAL) as leukaemias with one blast population scoring higher than 2 points in their scoring system for more than one lineage, and bilineal leukaemia if markers are expressed on distinct blast populations [13]. In contrast, ALs that do not meet lineage categorisation

Table 1. European Group for Immunological Characterization of Acute Leukaemias (EGIL), from Bene *et al.* [10]. Biphenotypic acute leukaemia (BAL) is diagnosed when scores are greater than 2 in 2 lineage columns. c, cytoplasmic; IgM, immunoglobulin M; MPO, myeloperoxidase; TCR, T-cell receptor; TdT, terminal deoxynucleotidyl transferase.

Points	B	T	Myeloid
2	cCD79a cCD22 cIgM	CD3 (surface or cytoplasmic) TCRab TCRgd	MPO
1	CD19 CD20 CD10	CD2 CD5 CD8 CD10	CD117 CD13 CD33 CD65w
0.5	TdT CD24	TdT CD7 Cd1a	CD14 CD15 CD64

Table 2. World Health Organization 2008/2016 Criteria, from Borowitz *et al.* [12] and Arber *et al.* [9]. MPO, myeloperoxidase, NSE, nonspecific esterase.

Lineage	Markers
Myeloid	MPO (flow cytometry, immunohistochemistry, or enzyme cytochemistry) -OR- Monocytic differentiation (at least 2 of the following: NSE cytochemistry, CD11c, CD14, CD64, lysozyme)
T lineage	Strong ^a cytoplasmic CD3 -OR- Surface CD3
B lineage	Strong ^a CD19 with at least 1 of the following strongly expressed: CD79a, cytoplasmic CD22, or CD10 -OR- Weak CD19 with at least 2 of the following strongly expressed: CD79a, cytoplasmic CD22, or CD10

^aStrong: at least as intense as in normal B or T cells.

criteria are termed acute leukaemia of ambiguous lineage (ALAL) by the WHO, with no distinction for biphenotypic or bilineal expression patterns [9,12,16]. The ALAL subgroup comprises acute undifferentiated leukaemias (AUL) that lack expression of any lineage-specific antigens, and mixed-phenotype acute leukaemias (MPAL) that express specific markers for more than one cellular lineage. Although AUL have no lineage-specific markers, they commonly express CD34, HLA-DR, CD38 and/or TdT and may therefore be difficult to distinguish from the diagnostic category of AML with minimal differentiation (equivalent to M0-AML FAB subgroup). While true AULs are extremely rare, MPALs comprise 2–5% of AL [17,18] and are further subdivided by cytogenetic abnormality, specifically t(9;22)/*BCR-ABL1* and t(v;11q23)/*KMT2A* (previously *MLL*, see below). Most cases do not harbour these alterations, so are classed as MPAL not otherwise specified (NOS) [9]. MPALs are much more frequently T/myeloid or B/myeloid than B/T or B/T/myeloid, which are both vanishingly rare [19]. These relative incidences are in keeping with modern concepts of multilineage competency in late lymphoid precursors (T/myeloid and B/myeloid) [1–8], whereas the existence of progenitors with exclusively B-/T-lymphoid oligopotency has not been proven.

Even in more ‘straightforward’ cases, categorisation by MFC is still affected by technical factors, observer interpretation and choice of diagnostic criteria. For example, retrospective analyses of both paediatric and adult cohorts have revealed higher incidences of EGIL-categorised BAL than WHO-categorised MPAL, which is almost certainly due to the more stringent criteria of the latter system [18,20–22]. Choice of diagnostic guidelines may therefore influence treatment allocation and inclusion on clinical studies, in turn introducing a bias in prognostic and genetic studies and therapy evaluation. Diagnosis may be further influenced by technical aspects of immunophenotyping, including sample quality, permeabilisation procedure, antibody clones, compensation and gating strategies. The final interpretation of antigen expression can also be heterogeneous, with positivity for some markers (especially MPO) being more subjective than others. Efforts have been made to reduce this variability by standardisation of MFC protocols and reagent panels, quality control and file sharing between expert groups [10,23–25]. For example, the EuroFlow consortium has developed antibody panels [24] and standardised protocols [25] for diagnosis of haematological malignancies. More recently, these efforts have extended to development of automated gating strategies [26,27] that would further improve diagnostic reproducibility

through more reliable and robust identification and characterisation of leukaemic blasts and normal blood cells. The AIEOP-BFM consortium has also recently provided recommendations for pragmatic thresholds for interpretation of immunophenotypic marker expression that included a consensus antigen panel designed to fulfil EGIL and WHO 2008/2016 requirements for AL subtyping [28]. Recent technological advances in cytometer machines, fluorochromes and informatics tools are also likely to allow better analysis of the phenotypic complexity of the leukemic population, which should also lead to more robust diagnostic practice in the future.

The molecular landscape of T/Myeloid leukaemias

Given that current MFC scoring approaches are not ideal for all, could assessment of other aspects of leukaemia biology aid diagnosis and treatment choice? ALs typically harbour mutations, deletions or translocations in factors that regulate normal blood cell development. Some of these molecules (usually transcription factors) have lineage-specific (or even precursor-specific) activities. We can therefore imagine that mutational genotype could earmark leukaemias with shared molecular mechanisms of dysregulation of differentiation processes, even in ALs that are phenotypically heterogeneous.

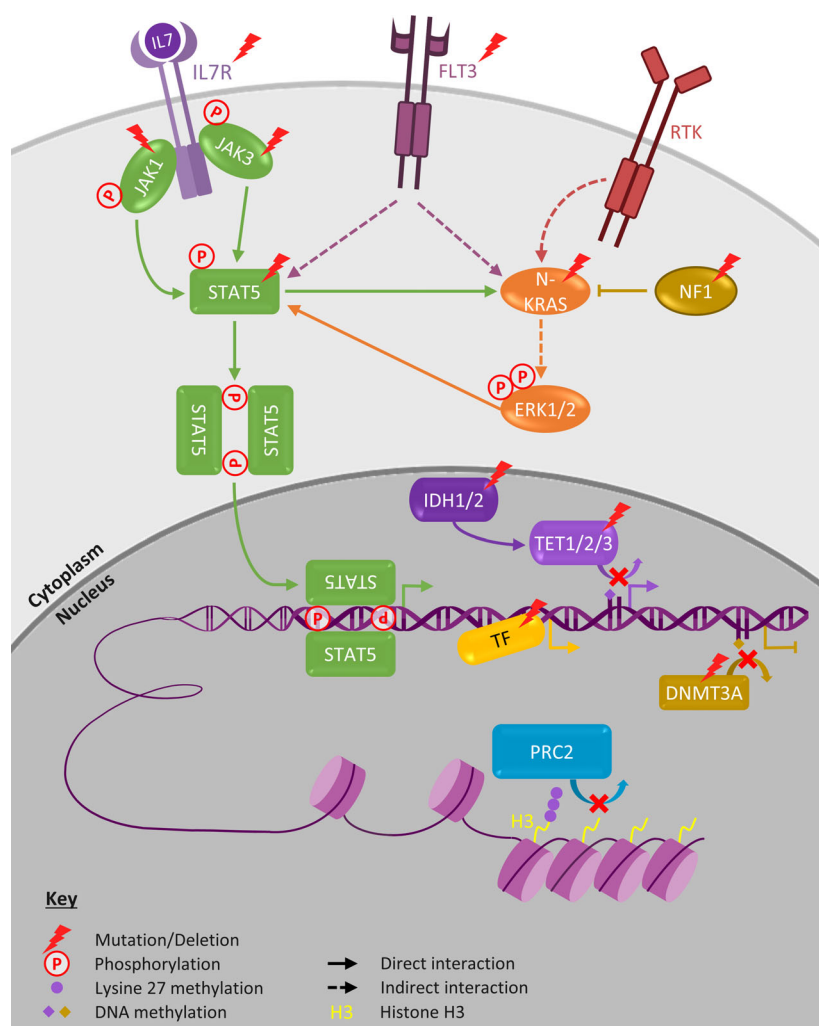
As we have seen, co-expression of T-lymphoid and myeloid antigens on AL blasts is not unusual. T/myeloid MPALs may be difficult to distinguish from early T-cell precursor (ETP)-ALL, which is defined phenotypically by absence of T-cell markers CD1a and CD8, weak or absent expression of CD5, and positivity for at least one haematopoietic stem cell and/or myeloid antigen (CD34, CD117, HLA-DR, CD13, CD33, CD11b, CD65). This subgroup was originally identified in paediatric cases by transcriptional proximity to normal murine ETPs, and frequent treatment resistance [29], and accounts for 12–15% of childhood T-ALL and 20–25% of adult cases [30,31]. In addition, the most phenotypically immature AML subgroup, M0-AML, often co-expresses lymphoid-associated antigens such as CD7 or TdT [32]. Distinction between these three diagnostic categories may therefore rely purely on determination of cytoplasmic CD3 and MPO positivity. To be provocative, is this immunophenotypic nit-picking useful? This is an important clinical question, as immature T-ALLs and AMLs are more chemoresistant and have poorer outcomes than more mature leukaemia subgroups [29,31,33–35].

By considering the normal haematopoietic framework (Fig. 1), we can imagine that these cases could all potentially arise from transformation of similar multi- or bipotent lympho-myeloid progenitors. Indeed, it has been proposed that T/myeloid MPAL, ETP-ALL and cases of M0-AML with evidence of T lineage identity should be defined as a single diagnostic entity of acute myeloid/T-lymphoblastic leukaemia (AMTL) [36]. This argument is supported by the shared repertoire of genetic alterations observed in these leukaemias, which has been extensively catalogued in recent years [17,30,31,37–42]. Selected examples of the molecular pathways affected by these mutations (shown diagrammatically in Fig. 2) will follow.

As might be expected, mutations and deletions in *transcription factors* that normally regulate T-lymphoid and myeloid gene expression programmes are commonly found in these AL subgroups. Notably,

paediatric T/Myeloid MPALs were reported to harbour alterations in transcription factor genes in 100% of cases in a recent study [17]. The high frequencies of *RUNX1* inactivating mutations targeting RUNT and transactivation (TAD) domains seen in each of ETP-ALL [30,31,37,43], AML-M0 [41,42,44] and T/myeloid MPAL [17,38] reflect the key role that the core binding factor transcription factor complex plays at diverse stages of myeloid and lymphoid differentiation [45,46]. Similar patterns are observed for alterations in genes coding for *ETV6*, with most mutations located in the PNT and ETS domain coding regions [37,38,47–49], and *WT1* mutations in the zinc finger and WT1 domains [17,40,42,44,49,50]. In line with its diverse and critical roles in T-cell development [51], *GATA3* loss-of-function mutations segregate with immature cases of T-ALL [30,31] and are found in T/Myeloid, but not B/Myeloid MPAL [38].

Fig. 2. Genetic alterations in immature T-lymphoid/myeloid leukaemias. Schematic depiction of molecular categories that are frequently mutated or deleted in ETP-ALL, T/Myeloid MPAL and M0-AML. Selected examples within the key functional categories (signalling molecules, epigenetic proteins and transcription factors) are shown. DNMT3A, DNA methyltransferase 3 alpha; ERK, extracellular signal-regulated kinase; FLT3, Fms-related receptor tyrosine kinase 3; H3, histone H3; IDH, isocitrate dehydrogenase; IL7R, interleukin 7 receptor; JAK, Janus kinase; NF1, neurofibromin 1; PRC2, polycomb repressive complex 2; RTK, receptor tyrosine kinase; STAT, signal transducer and activator of transcription; TET, Tet methylcytosine dioxygenase; TF, transcription factor.



Signalling molecule mutations are also frequent in these leukaemia subgroups. T/Myeloid MPALs and ETP-ALLs harbour particularly high rates of activating mutations in members of the IL7R-JAK-STAT phosphorylation cascade, causing marked effects on pathway activity and disease biology [52–54]. Kinase gene alterations that link to aggressive AML biology, such as *FLT3* activation by either mutation or internal tandem duplication [37,42,55] and *NRAS* gain of function mutations (most frequently at codon 12), are also common in these cases [17,30,31,37,39,40,42,50,55–60]. This is in contrast to patterns of signalling alteration seen in more phenotypically mature leukaemias. For example, NOTCH pathway hyperactivity due to either *NOTCH1* activating mutations or *FBXW7* inactivating perturbations is more common in cases of mature T-ALL, albeit high-throughput sequencing has revealed higher incidences of these mutations in ETP-ALL than were previously recognised [17,30,31,58,61]. Similarly, PI3K-AKT-mTOR signalling deregulation, often due to *PTEN* inactivating mutations that may link to altered leukaemia stem cell activity [62], is more frequent in mature T-ALL subgroups [63,64]. There is additional evidence of interplay between cellular signalling pathways and transcription factor action in AL. As an example, the activity of MEF2C, which is highly expressed in some cases of immature T-ALL [65], has been shown to be altered both directly and indirectly by MAPK and SIK kinases in a distinct subset of AML [66,67].

An association between lineage-ambiguous leukaemias and *epigenetic dysfunction* has long been established, since the initial descriptions of pathogenic translocations in the histone methyltransferase gene *KMT2A* (previously *MLL*, or mixed lineage leukaemia [68,69]), which are highly prevalent in infant AL [70]. As epigenetic factors are key regulators of normal haematopoiesis [71–74], it is to be expected that the molecular dysregulation caused by genetic alterations in these factors would be leukaemogenic. It is now evident that mutations and deletions in factors that chemically modify either DNA or histones are common in blood cancers [75–78] and that epigenetic alterations are linked to marked deregulation of gene expression in leukaemic cells [79–81].

Several studies have highlighted increased rates of Polycomb factor mutations and deletions in ETP-ALL compared with more mature T-ALL subgroups [30,31,57], and these are also found in T/Myeloid MPAL and M0-AML [17,30,38,39,42]. Specifically, these loss-of-function alterations usually occur in components of the polycomb repressive complex 2 (PRC2) that critically regulates developmental gene expression,

including in the haematopoietic system [73,74], through methylation of a specific histone lysine residue (H3K27) that is normally associated with transcriptional repression [82–85]. In the context of T-ALL, loss of PRC2 function due to mutation has been linked to altered K27 methylation that correlates with increased NOTCH pathway activity in leukaemic cells [86]. In line with this, PRC2 has recently been shown to play a key role in the delineation of the chromatin landscape during normal thymopoiesis [87]. We recently showed that haploinsufficiency of core PRC2 components EZH2, EED and SUZ12 due to mutation or deletion was strongly linked to poor outcome in childhood AML [88]. Interestingly, PRC2-depleted AMLs in our series were more likely to show weak MPO staining and/or express CD7, suggesting that some of these cases may show at least some phenotypic proximity to T/Myeloid MPALs.

Factors that control the methylation state of DNA are also often subject to leukaemia-associated dysregulation. Mutations in the *de novo* DNA methylating factor *DNMT3A* have been detected across the T-lymphoid and myeloid leukaemic spectrum [17,38,39,42,49,50,57,89] and are associated with poor prognosis in both AML [90,91] and T-ALL [43,92]. *DNMT3A* alterations are linked to increasing age at disease presentation and probably arise on a background of clonal haematopoiesis in most cases [93,94]. In contrast, mutations in isocitrate dehydrogenase genes (*IDH*) 1 and 2, which affect DNA methylation through generation of the oncometabolite 2-hydroxyglutarate [95], are found across all age groups [38–40,42,50] and are found more commonly in M0-AML than in more mature AML subtypes [42].

As these epigenetic factors act across the genome, alteration of their activity will result in widespread and pleiotropic effects on AL transcription. Furthermore, it is likely that the leukaemic epigenetic ‘state’ is also heavily influenced by cellular lineage and precursor type, in the absence and presence of somatic mutations. As an example, DNA methylation patterns were reported to be highly divergent between T/Myeloid and B/Myeloid MPAL, in some cases overlapping with more differentiated ALs [38]. While T/Myeloid MPALs showed higher levels of methylation in general, T-lymphoid ontogeny was reflected by hypomethylation of T-receptor signalling factor genes. Also in line with these concepts of global epigenetic dysregulation, recent intriguing data [96] have suggested that some cases of AML have epigenomic landscape changes reminiscent of the MYB regulatory complex reorganisation that is also found in T-ALL [97,98].

The genome-wide effects of epigenetic factor alterations are likely to be further affected by co-occurring mutations. Some clues may be seen in the case of PHF6, a chromatin-binding protein that modulates the recruitment of lineage-specific transcription factors and chromatin-remodelling agents [99]. *PHF6* is widely mutated in AL, with a strong predominance in T-ALL, where loss-of-function mutations are found in 16% of paediatric and 38% of adult cases [100]. The majority of *PHF6*-mutated T-ALLs show immature surface phenotypes, and analysis of clonal evolution and mutation dynamics using diagnostic and relapse samples identified these mutations as early events in T-ALL development [100]. Mutations are also found in 3% of AML, but in contrast to T-ALL, a majority are acquired in subclonal populations in the later phase of the clinical course [101]. Approximately 16–55% of mixed-phenotype acute leukaemia also harbour these mutations [102], particularly in the very rare T-B-MPAL group [49]. No direct functional evidence has been found to demonstrate whether *PHF6* mutations contribute to pathogenesis, albeit there have been reports that PHF6 affects lineage plasticity [103] and that it may regulate expression of the key lymphoid transcription factor *LMO2* [104]. Murine *PHF6* knockout models develop haematopoietic neoplasms with various phenotypes after a long latency, demonstrating a tumour-suppressive function that may be linked to reduced thresholds for oncogenic transformation and increased frequency of leukaemia-initiating cells [49]. This long latency also suggests that PHF6 loss is not sufficient to induce overt leukaemia, and that cooperative mutations that are required for malignant transformation, and may also influence leukaemia phenotype. In support of this concept, *PHF6* mutations are seen alongside *NOTCH1* alterations in T-ALL [100,105,106], but not in other AL types. We have also found that *PHF6*-mutated AL transcriptional profiles might segregate with *NOTCH1* genotype [107], raising the possibility that co-occurring mutations affect leukaemia phenotype by altering lineage-specific gene expression programmes. It is to be hoped that ongoing research into the role of PHF6 in leukaemia and normal haematopoiesis will shed light on these complex effects.

New technologies, new perspectives?

As well as providing a thorough description of leukaemia-associated mutations and deletions, next-generation sequencing technologies have generated extensive information about the transcriptional and epigenetic landscape of AL, which can offer further useful

insights into the molecular roadmap of immature acute leukaemias.

It is well-established that gene expression profiling can be used to subcategorise cancers, including leukaemias according to shared transcriptional signatures [108,109]. Given that AL-related mutations and deletions tend to recurrently converge on specific molecular pathways that are often linked to haematopoietic differentiation, it can be predicted that leukaemic gene expression profiles might cluster according to the underlying mechanism of oncogenesis. In line with this, several independent studies have shown that unsupervised *hierarchical clustering* of T-ALL transcriptional profiles reproducibly group cases according to a limited number of expression signatures [65,110,111]. These clusters correlate closely with the stage of T-lymphoid differentiation arrest, and so typically identify a distinct immature/ETP-ALL subgroup. Other tools like *gene set enrichment analysis (GSEA)* compare enrichment for defined transcriptional signatures between different sample groups, so can therefore be used to measure the relative activity of specific molecular programmes, or to assess transcriptional resemblance to other cell types [112]. In the case of ETP-ALLs, GSEA has identified shared gene expression programmes with both normal haematopoietic stem cell (HSC) and immature myeloid precursors [37], thereby providing clues about the leukaemia cell of origin, and the potential molecular basis of the differentiation block. However, most traditional analytical techniques are hampered by several factors, including reliance on comparisons of predefined sample groups, and limited capacity to either account for transcriptional heterogeneity between individual leukaemias, or to interpret and resolve relationships between the subgroups that are identified.

Once again, we should consider how analysis of normal haematopoiesis might inform our understanding of leukaemia. High-throughput single-cell sequencing approaches have yielded many important insights into normal blood cell development in recent years [113,114]. These technologies have stimulated the development of a range of new analytical and visualisation tools to help unravel the complex data they generate, and some of these methods have already been applied to AL data sets. One example is *stochastic neighbour embedding (SNE)*. As with other dimensionality reduction techniques such as Principal Component Analysis (PCA), SNE reduces complex information to a small number of dimensions, so can therefore help to visualise clusters of biologically similar samples in large multidimensional data sets, in a way that generally agrees with dedicated clustering algorithms [115]. SNE has been

used for single-cell transcriptomic analysis [116], but can also be applied to leukaemia bulk gene expression profiles [117–122]. For example, t-SNE was used together with predictive modelling to help discover novel subgroups of paediatric B-ALL linked by dysregulation of the B-lymphoid transcription factor *PAX5* [120]. In the context of immature T/Myeloid leukaemias, an analysis of ETP-ALLs MPALs and AMLs was reported to identify molecular subgroups distinguished by the presence of either *FLT3* or *PRC2* alterations, which were associated with differences in patient prognosis [119]. Recent emerging data showed that t-SNE helped to identify a novel immature AL subset characterised by enhancer hijacking of the T-lymphoid regulator *BCL11B* [122]. Overall, these data suggest that SNE can help to reveal biological relationships between ALs that could provide novel insights into leukaemogenesis.

Iterative Clustering and Guide Gene Selection (ICGS) provides another example of a tool that was initially developed for single-cell transcriptional analysis [123], but which might also be used for leukaemia profiling.

Briefly, ICGS relies on clustering samples using a limited number of ‘guide genes’ that are sufficient to define more extensive gene expression patterns, including lineage-related programmes. Serial clustering and refinement of these guide genes allow case grouping according to shared transcriptional modules, thereby identifying common expression signatures between sample subsets. Importantly, analysis of haematopoietic progenitors has shown that ICGS can infer ontogeny relationships between cellular developmental states [123]. We recently used ICGS to analyse the transcriptional profiles of a large cohort of ALs that included a high proportion of immature T-lymphoid and myeloid leukaemias [107]. In contrast to traditional clustering methods, we found that ICGS could reveal a continuum of haematopoietic differentiation between T-lymphoid and myeloid cases, with a significant subset of cases grouping together at the T-lymphoid/myeloid interface (Fig. 3). We used GSEA to show that these ‘interface’ acute leukaemias (IALs) shared gene expression programmes with a variety of multi- or bipotent myeloid and lymphoid hematopoietic

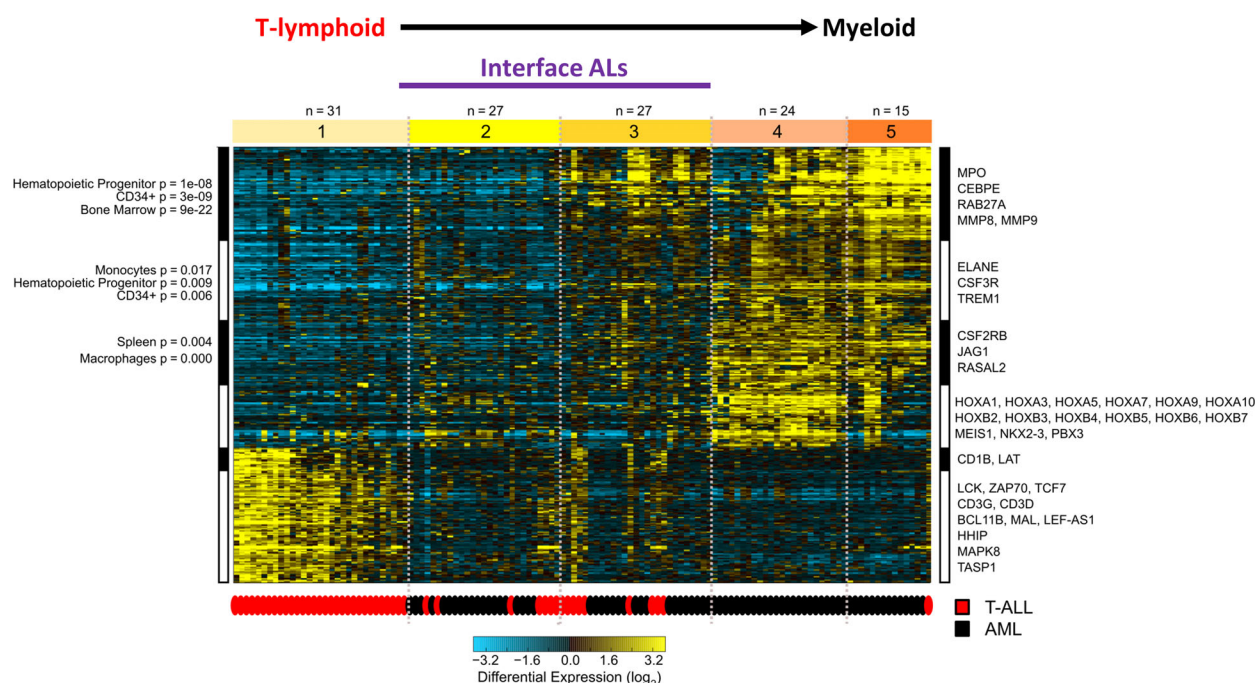


Fig. 3. Iterative Clustering and Guide Gene Selection (ICGS). Heatmap showing expression of guide genes selected by ICGS during a transcriptional analysis of 48 T-ALL and 76 AML samples. Columns represent individual cases, with the original diagnostic category (T-ALL or AML) denoted along the bottom. Bars along the top indicate the 5 clusters identified by ICGS, corresponding to a continuum of differentiation between T-lymphoid and myeloid cases. Clusters 2 and 3 were designated as interface acute leukaemias (IALs), as discussed in the text. Rows represent genes, and bars on the side correspond to blocks of correlated genes. Selected enriched gene ontology groups are shown. This image is modified from the original publication of these data (Bond *et al*, Leukemia 2020 [107]) and is reproduced under the terms of a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

precursor populations, including the most immature and rare CD34+CD1a-CD7- subset of human ETPs. Surprisingly, we found that IALs originally diagnosed as AML had transcriptional resemblance to lymphoid progenitor populations by GSEA, raising the possibility that these cases harbour latent lymphoid identity that cannot be detected by current diagnostic immunophenotyping panels. We further found that the expression of IAL transcriptional programmes correlated with poor prognosis in independent AML cohorts [124,125], suggesting that myeloid-directed therapies might not provide optimal treatment for these cases. Of note, while ETP-ALLs and M0-AMLs were more likely to cluster at the lineage interface in our analyses, IALs were not limited to these cases, and somatic mutational genotypes showed only partial correlation with known alteration patterns in immature ALs.

As well as transcriptional profiling, it is now possible to use high-throughput approaches to investigate many other biological parameters at single-cell resolution, thereby creating opportunities to perform a highly granular and holistic analysis of transcriptional, epigenetic and phenotypic relationships in both normal and leukaemia data sets. As an example, a recent study used CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing [126]) and ATAC-seq (assay for transposase-accessible chromatin using sequencing [127]) to analyse normal haematopoietic precursors and MPALs at a single-cell level [128]. Crucially, integration of single-cell immunophenotypic, transcriptional and epigenetic data allowed generation of a haematopoietic roadmap without specific boundaries between cell populations, thereby recapitulating the normal continuum of differentiation. The latter was visualised by *UMAP* [129], a dimensionality reduction method similar to *t-SNE*, but which better preserves potential hierarchical structures within the data set, and which is also well suited for projection of distinct independently acquired data. By projecting MPAL profiles onto this roadmap, the authors could interrogate differences in transcriptional and epigenetic programmes between leukaemia cells and the corresponding normal precursor, allowing them to identify a specific disruption in *RUNX1*-directed gene expression in AL samples. This transcriptional signature was further correlated with poor outcomes in patients with AML, providing a direct link between this very specific molecular dysfunction and aggressive leukaemia biology [128].

Conclusions

Improvements in AL outcomes, particularly in children, have been one of the success stories of modern

medicine. These advances have been underpinned by careful dissection of leukaemia molecular biology that has developed with reference to normal blood cell development. While current classification criteria allow many patients to receive effective treatments, new haematopoietic paradigms should lead us to consider whether these systems should be modified for ALs that seem to straddle the lineage spectrum.

In contrast to categorical approaches, we have seen that ICGS, and other analytical methods for single cell -omics analysis can identify similarities and deviations from the normal haematopoietic continuum in human leukaemias, which in turn could be used to discover novel molecular pathology. We hope that this knowledge could inform AL diagnosis in practice, either by incorporation of new molecular assessments, or through refinement of MFC approaches based on the information generated by these novel tools. In the case of immature T/myeloid leukaemias, there is already persuasive evidence that rational treatment strategies might target common molecular alterations in cases that are currently categorised heterogeneously in clinical practice.

In summary, we believe that ongoing reference to an up-to-date haematopoietic roadmap will help to provide a better diagnostic and treatment journey for patients with acute leukaemia.

Acknowledgements

Work in the Bond laboratory in Dublin was supported by a National Children's Research Centre Leadership Award (NCRC A/18/3). LJ was also supported by an Irish Cancer Society Research Fellowship (CRF20JON). Work in the Laurenti laboratory in Cambridge was supported by a Wellcome/Royal Society Sir Henry Dale Fellowship to EL, the European Haematology Association, BBSRC and by core funding from Wellcome and MRC to the Wellcome-MRC Cambridge Stem Cell Institute. Work in the Macintyre laboratory in Paris was supported by the Association Laurette Fugain, La ligue contre le Cancer and the INCa 2007 (DM/FC/sl/RT07) and CAMELE Translational Research and PhD programmes and the INCa/AP-HP genetic Plateforme de Ressources Biologiques (PRB).

Conflict of interest

The authors declare no conflict of interest.

Authors contributions

All authors contributed to the writing of the manuscript.

References

- 1 Belluschi S, Calderbank EF, Ciaurro V, Pijuan-Sala B, Santoro A, Mende N, Diamanti E, Sham KYC, Wang X, Lau WWY *et al.* (2018) Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat Commun* **9**, 4100.
- 2 Doulatov S, Notta F, Laurenti E & Dick JE (2012) Hematopoiesis: a human perspective. *Cell Stem Cell* **10**, 120–136.
- 3 Goardon N, Marchi E, Atzberger A, Quek L, Schuh A, Soneji S, Woll P, Mead A, Alford KA, Rout R *et al.* (2011) Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138–152.
- 4 Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, Quek L, Otto G, Repapi E, Doondea J *et al.* (2018) Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol* **19**, 85–97.
- 5 Alhaj Hussien K, Vu Manh TP, Guimiot F, Nelson E, Chabaane E, Delord M, Barbier M, Berthault C, Dulphy N, Alberdi AJ *et al.* (2017) Molecular and functional characterization of lymphoid progenitor subsets reveals a bipartite architecture of human lymphopoiesis. *Immunity* **47**, 680–696.e8.
- 6 Kohn LA, Hao QL, Sasidharan R, Parekh C, Ge S, Zhu Y, Mikkola HK & Crooks GM (2012) Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin. *Nat Immunol* **13**, 963–971.
- 7 Hao QL, George AA, Zhu J, Barsky L, Zielinska E, Wang X, Price M, Ge S & Crooks GM (2008) Human intrathymic lineage commitment is marked by differential CD7 expression: identification of CD7- lympho-myeloid thymic progenitors. *Blood* **111**, 1318–1326.
- 8 Luc S, Luis TC, Boukarabila H, Macaulay IC, Buza-Vidas N, Bouriez-Jones T, Lutteropp M, Woll PS, Loughran SJ, Mead AJ *et al.* (2012) The earliest thymic T cell progenitors sustain B cell and myeloid lineage potential. *Nat Immunol* **13**, 412–419.
- 9 Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M & Vardiman JW (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405.
- 10 Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A & van't Veer MB (1995) Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). *Leukemia* **9**, 1783–1786.
- 11 Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellstrom-Lindberg E, Tefferi A *et al.* (2009) The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937–951.
- 12 Borowitz MJ, Bene MC, Harris NL, Porwit A & Matutes E (2008) Acute leukaemias of ambiguous lineage. In WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H & Thiele J eds), pp. 179–187. International Agency for Research on Cancer, Lyon.
- 13 Rubnitz JE, Onciu M, Pounds S, Shurtleff S, Cao X, Raimondi SC, Behm FG, Campana D, Razzouk BI, Ribeiro RC *et al.* (2009) Acute mixed lineage leukemia in children: the experience of St Jude Children's Research Hospital. *Blood* **113**, 5083–5089.
- 14 Kurosawa S, Taya T, Kishida Y, Nagata A, Yamada Y, Konishi T, Kaito S, Yoshifuji K, Matsuyama N, Shirane S *et al.* (2018) Outcome of patients with acute undifferentiated leukemia after allogeneic hematopoietic stem cell transplantation. *Leuk Lymphoma* **59**, 3006–3009.
- 15 Wolach O & Stone RM (2015) How I treat mixed-phenotype acute leukemia. *Blood* **125**, 2477–2485.
- 16 Bene MC & Porwit A (2012) Acute leukemias of ambiguous lineage. *Semin Diagn Pathol* **29**, 12–18.
- 17 Alexander TB, Gu Z, Iacobucci I, Dickerson K, Choi JK, Xu B, Payne-Turner D, Yoshihara H, Loh ML, Horan J *et al.* (2018) The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature* **562**, 373–379.
- 18 Weinberg OK & Arber DA (2010) Mixed-phenotype acute leukemia: historical overview and a new definition. *Leukemia* **24**, 1844–1851.
- 19 Matutes E, Pickl WF, Van't Veer M, Morilla R, Swansbury J, Strobl H, Attarbaschi A, Hopfinger G, Ashley S, Bene MC *et al.* (2011) Mixed-phenotype acute leukemia: clinical and laboratory features and outcome in 100 patients defined according to the WHO 2008 classification. *Blood* **117**, 3163–3171.
- 20 Yan L, Ping N, Zhu M, Sun A, Xue Y, Ruan C, Drexler HG, Macleod RA, Wu D & Chen S (2012) Clinical, immunophenotypic, cytogenetic, and molecular genetic features in 117 adult patients with mixed-phenotype acute leukemia defined by WHO-2008 classification. *Haematologica* **97**, 1708–1712.
- 21 van den Ancker W, Terwijn M, Westers TM, Merle PA, van Beckhoven E, Drager AM, Ossenkoppele GJ & van de Loosdrecht AA (2010) Acute leukemias of ambiguous lineage: diagnostic consequences of the WHO2008 classification. *Leukemia* **24**, 1392–1396.
- 22 Lee HG, Baek HJ, Kim HS, Park SM, Hwang TJ & Kook H (2019) Biphenotypic acute leukemia or acute

- leukemia of ambiguous lineage in childhood: clinical characteristics and outcome. *Blood Res* **54**, 63–73.
- 23 van der Does-Van den Berg A, Bartram CR, Basso G, Benoit YCM, Biondi A, Debatin K-M, Haas OA, Harbott J, Kamps WA, Köller U *et al.* (1992) Minimal requirements for the diagnosis, classification, and evaluation of the treatment of childhood acute lymphoblastic leukemia (ALL) in the "BFM Family" Cooperative Group. *Med Pediatr Oncol* **20**, 497–505.
 - 24 van Dongen JJ, Lhermitte L, Bottcher S, Almeida J, van der Velden VH, Flores-Montero J, Rawstron A, Asnafi V, Lecrevisse Q, Lucio P *et al.* (2012) EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia* **26**, 1908–1975.
 - 25 Kalina T, Flores-Montero J, van der Velden VH, Martin-Ayuso M, Bottcher S, Ritgen M, Almeida J, Lhermitte L, Asnafi V, Mendonca A *et al.* (2012) EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia* **26**, 1986–2010.
 - 26 Lhermitte L, Mejstrikova E, van der Sluijs-Gelling AJ, Grigore GE, Sedek L, Bras AE, Gaipa G, Sobral da Costa E, Novakova M, Sonneveld E *et al.* (2018) Automated database-guided expert-supervised orientation for immunophenotypic diagnosis and classification of acute leukemia. *Leukemia* **32**, 874–881.
 - 27 Lhermitte L, Barreau S, Morf D, Fernandez P, Grigore G, Barrena S, de Bie M, Flores-Montero J, Bruggemann M, Mejstrikova E *et al.* (2020) Automated identification of leukocyte subsets improves standardization of database-guided expert-supervised diagnostic orientation in acute leukemia: a EuroFlow study. *Mod Pathol* **34**, 59–69.
 - 28 Dworzak MN, Buldini B, Gaipa G, Ratei R, Hrusak O, Luria D, Rosenthal E, Bourquin JP, Sartor M, Schumich A *et al.* (2018) AIEOP-BFM consensus guidelines 2016 for flow cytometric immunophenotyping of Pediatric acute lymphoblastic leukemia. *Cytometry B Clin Cytom* **94**, 82–93.
 - 29 Coustan-Smith E, Mullighan CG, Onciu M, Behm FG, Raimondi SC, Pei D, Cheng C, Su X, Rubnitz JE, Basso G *et al.* (2009) Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol* **10**, 147–156.
 - 30 Liu Y, Easton J, Shao Y, Maciaszek J, Wang Z, Wilkinson MR, McCastlain K, Edmonson M, Pounds SB, Shi L *et al.* (2017) The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat Genet* **49**, 1211–1218.
 - 31 Bond J, Graux C, Lhermitte L, Lara D, Cluzeau T, Leguay T, Cieslak A, Trinquand A, Pastoret C, Belhocine M *et al.* (2017) Early response-based therapy stratification improves survival in adult early thymic precursor acute lymphoblastic leukemia: a group for research on adult acute lymphoblastic leukemia study. *J Clin Oncol* **35**, 2683–2691.
 - 32 Gougounon A, Abahssain H, Rigollet L, Elhamri M, Tigaud I, Chelghoum Y, Plesa A, Dumontet C, Michallet M & Thomas X (2011) Minimally differentiated acute myeloid leukemia (FAB AML-M0): prognostic factors and treatment effects on survival—a retrospective study of 42 adult cases. *Leuk Res* **35**, 1027–1031.
 - 33 Wood BL, Winter SS, Dunsmore KP, Devidas M, Chen S, Asselin B, Esiashvili N, Loh ML, Winick NJ, Carroll WL *et al.* (2014) T-lymphoblastic leukemia (T-ALL) shows excellent outcome, lack of significance of the early thymic precursor (ETP) immunophenotype, and validation of the prognostic value of end-induction minimal residual disease (MRD) in Children's Oncology Group (COG) study AALL0434. *Blood* **124**, 1.
 - 34 Walter RB, Othus M, Burnett AK, Lowenberg B, Kantarjian HM, Ossenkoppele GJ, Hills RK, van Montfort KG, Ravandi F, Evans A *et al.* (2013) Significance of FAB subclassification of "acute myeloid leukemia, NOS" in the 2008 WHO classification: analysis of 5848 newly diagnosed patients. *Blood* **121**, 2424–2431.
 - 35 Barbaric D, Alonzo TA, Gerbing RB, Meshinchi S, Heerema NA, Barnard DR, Lange BJ, Woods WG, Arceci RJ & Smith FO (2007) Minimally differentiated acute myeloid leukemia (FAB AML-M0) is associated with an adverse outcome in children: a report from the Children's Oncology Group, studies CCG-2891 and CCG-2961. *Blood* **109**, 2314–2321.
 - 36 Gutierrez A & Kentsis A (2018) Acute myeloid/T-lymphoblastic leukaemia (AMTL): a distinct category of acute leukaemias with common pathogenesis in need of improved therapy. *Br J Haematol* **180**, 919–924.
 - 37 Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M *et al.* (2012) The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163.
 - 38 Takahashi K, Wang F, Morita K, Yan Y, Hu P, Zhao P, Zhar AA, Wu CJ, Gumbs C, Little L *et al.* (2018) Integrative genomic analysis of adult mixed phenotype acute leukemia delineates lineage associated molecular subtypes. *Nat Commun* **9**, 2670.
 - 39 Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N *et al.* (2016) Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med* **374**, 2209–2221.
 - 40 Bolouri H, Farrar JE, Triche T Jr, Ries RE, Lim EL, Alonzo TA, Ma Y, Moore R, Mungall AJ, Marra MA *et al.* (2018) The molecular landscape of pediatric

- acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* **24**, 103–112.
- 41 Rose D, Haferlach T, Schnittger S, Perglerova K, Kern W & Haferlach C (2017) Subtype-specific patterns of molecular mutations in acute myeloid leukemia. *Leukemia* **31**, 11–17.
 - 42 Kao HW, Liang DC, Wu JH, Kuo MC, Wang PN, Yang CP, Shih YS, Lin TH, Huang YH & Shih LY (2014) Gene mutation patterns in patients with minimally differentiated acute myeloid leukemia. *Neoplasia* **16**, 481–488.
 - 43 Grossmann V, Haferlach C, Weissmann S, Roller A, Schindela S, Poetzinger F, Stadler K, Bellos F, Kern W, Haferlach T *et al.* (2013) The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. *Genes Chromosomes Cancer* **52**, 410–422.
 - 44 Lin PH, Li HY, Fan SC, Yuan TH, Chen M, Hsu YH, Yang YH, Li LY, Yeh SP, Bai LY *et al.* (2017) A targeted next-generation sequencing in the molecular risk stratification of adult acute myeloid leukemia: implications for clinical practice. *Cancer Med* **6**, 349–360.
 - 45 Mevel R, Draper JE, Lie ALM, Kouskoff V & Lacaud G (2019) RUNX transcription factors: orchestrators of development. *Development* **146**.
 - 46 Shin B, Hosokawa H, Romero-Wolf M, Zhou W, Masuhara K, Tobin VR, Levanon D, Groner Y & Rothenberg EV (2021) Runx1 and Runx3 drive progenitor to T-lineage transcriptome conversion in mouse T cell commitment via dynamic genomic site switching. *Proc Natl Acad Sci USA* **118**, e2019655118.
 - 47 Van Vlierberghe P, Ambesi-Impiombato A, Perez-Garcia A, Haydu JE, Rigo I, Hadler M, Tosello V, Della Gatta G, Paietta E, Racevskis J *et al.* (2011) ETV6 mutations in early immature human T cell leukemias. *J Exp Med* **208**, 2571–2579.
 - 48 Hock H, Meade E, Medeiros S, Schindler JW, Valk PJ, Fujiwara Y & Orkin SH (2004) Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev* **18**, 2336–2341.
 - 49 Xiao W, Bharadwaj M, Levine M, Farnhoud N, Pastore F, Getta BM, Hultquist A, Famulare C, Medina JS, Patel MA *et al.* (2018) PHF6 and DNMT3A mutations are enriched in distinct subgroups of mixed phenotype acute leukemia with T-lineage differentiation. *Blood Adv* **2**, 3526–3539.
 - 50 Eckstein OS, Wang L, Punia JN, Kornblau SM, Andreeff M, Wheeler DA, Goodell MA & Rau RE (2016) Mixed-phenotype acute leukemia (MPAL) exhibits frequent mutations in DNMT3A and activated signaling genes. *Exp Hematol* **44**, 740–744.
 - 51 Rothenberg EV, Ungerback J & Champhekar A (2016) Forging T-lymphocyte identity: intersecting networks of transcriptional control. *Adv Immunol* **129**, 109–174.
 - 52 Vicente C, Schwab C, Broux M, Geerdens E, Degryse S, Demeyer S, Lahortiga I, Elliott A, Chilton L, La Starza R *et al.* (2015) Targeted sequencing identifies associations between IL7R-JAK mutations and epigenetic modulators in T-cell acute lymphoblastic leukemia. *Haematologica* **100**, 1301–1310.
 - 53 de Bock CE, Demeyer S, Degryse S, Verbeke D, Sweron B, Gielen O, Vandepoel R, Vicente C, Vanden Bempt M, Dagklis A *et al.* (2018) HOXA9 cooperates with activated JAK/STAT signaling to drive leukemia development. *Cancer Discov* **8**, 616–631.
 - 54 Kim R, Boissel N, Touzart A, Leguay T, Thonier F, Thomas X, Raffoux E, Huguet F, Villarese P, Fourrage C *et al.* (2020) Adult T-cell acute lymphoblastic leukemias with IL7R pathway mutations are slow-responders who do not benefit from allogeneic stem-cell transplantation. *Leukemia* **34**, 1730–1740.
 - 55 Noronha EP, Marques LVC, Andrade FG, Sardou-Cezar I, Dos Santos-Bueno FV, Zampier CDP, Terra-Granado E & Pombo-de-Oliveira MS (2019) T-lymphoid/myeloid mixed phenotype acute leukemia and early T-cell precursor lymphoblastic leukemia similarities with NOTCH1 mutation as a good prognostic factor. *Cancer Manag Res* **11**, 3933–3943.
 - 56 Neumann M, Heesch S, Gokbuget N, Schwartz S, Schlee C, Benlasfer O, Farhadi-Sartangi N, Thibaut J, Burmeister T, Hoelzer D *et al.* (2012) Clinical and molecular characterization of early T-cell precursor leukemia: a high-risk subgroup in adult T-ALL with a high frequency of FLT3 mutations. *Blood Cancer J* **2**, e55.
 - 57 Neumann M, Heesch S, Schlee C, Schwartz S, Gokbuget N, Hoelzer D, Konstandin NP, Ksienzyk B, Vosberg S, Graf A *et al.* (2013) Whole-exome sequencing in adult ETP-ALL reveals a high rate of DNMT3A mutations. *Blood* **121**, 4749–4752.
 - 58 Noronha EP, Marques LVC, Andrade FG, Thuler LCS, Terra-Granado E, Pombo-de-Oliveira MS, & Brazilian Collaborative Study Group of Acute, L (2019) The profile of immunophenotype and genotype aberrations in subsets of pediatric T-cell acute lymphoblastic leukemia. *Front Oncol* **9**, 316.
 - 59 Takahashi W, Sasaki K, Kvomatsu N & Mitani K (2005) TEL/ETV6 accelerates erythroid differentiation and inhibits megakaryocytic maturation in a human leukemia cell line UT-7/GM. *Cancer Sci* **96**, 340–348.
 - 60 Zhang X, Su J, Jeong M, Ko M, Huang Y, Park HJ, Guzman A, Lei Y, Huang YH, Rao A *et al.* (2016) DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet* **48**, 1014–1023.

- 61 Chang YH, Yu CH, Jou ST, Lin CY, Lin KH, Lu MY, Wu KH, Chang HH, Lin DT, Lin SW *et al.* (2021) Targeted sequencing to identify genetic alterations and prognostic markers in pediatric T-cell acute lymphoblastic leukemia. *Sci Rep* **11**, 769.
- 62 Guo W, Lasky JL, Chang CJ, Mosessian S, Lewis X, Xiao Y, Yeh JE, Chen JY, Iruela-Arispe ML, Varella-Garcia M *et al.* (2008) Multi-genetic events collaboratively contribute to Pten-null leukaemia stem-cell formation. *Nature* **453**, 529–533.
- 63 Tesio M, Trinquand A, Ballerini P, Hypolite G, Lhermitte L, Petit A, Ifrah N, Baruchel A, Dombret H, Macintyre E *et al.* (2017) Age-related clinical and biological features of PTEN abnormalities in T-cell acute lymphoblastic leukaemia. *Leukemia* **31**, 2594–2600.
- 64 Trinquand A, Tanguy-Schmidt A, Ben Abdelali R, Lambert J, Beldjord K, Lengline E, De Gunzburg N, Payet-Bornet D, Lhermitte L, Mossafa H *et al.* (2013) Toward a NOTCH1/FBXW7/RAS/PTEN-based oncogenetic risk classification of adult T-cell acute lymphoblastic leukemia: a Group for Research in Adult Acute Lymphoblastic Leukemia study. *J Clin Oncol* **31**, 4333–4342.
- 65 Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M, Vuerhard M, Buijs-Gladdines J, Kooi C, Klous P *et al.* (2011) Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* **19**, 484–497.
- 66 Brown FC, Still E, Koche RP, Yim CY, Takao S, Cifani P, Reed C, Gunasekera S, Ficarro SB, Romanienko P *et al.* (2018) MEF2C phosphorylation is required for chemotherapy resistance in acute myeloid leukemia. *Cancer Discov* **8**, 478–497.
- 67 Tarumoto Y, Lu B, Somerville TDD, Huang YH, Milazzo JP, Wu XS, Klingbeil O, El Demerdash O, Shi J & Vakoc CR (2018) LKB1, salt-inducible kinases, and MEF2C are linked dependencies in acute myeloid leukemia. *Mol Cell* **69**, 1017–1027.e6.
- 68 Gu Y, Nakamura T, Alder H, Prasad R, Canaani O, Cimino G, Croce CM & Canaani E (1992) The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. *Cell* **71**, 701–708.
- 69 Tkachuk DC, Kohler S & Cleary ML (1992) Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**, 691–700.
- 70 Meyer C, Burmeister T, Groger D, Tsaur G, Fechina L, Renneville A, Sutton R, Venn NC, Emerenciano M, Pombo-de-Oliveira MS *et al.* (2018) The MLL recombinome of acute leukemias in 2017. *Leukemia* **32**, 273–284.
- 71 Cedar H & Bergman Y (2011) Epigenetics of haematopoietic cell development. *Nat Rev Immunol* **11**, 478–488.
- 72 Gore AV & Weinstein BM (2016) DNA methylation in hematopoietic development and disease. *Exp Hematol* **44**, 783–790.
- 73 Xie H, Xu J, Hsu JH, Nguyen M, Fujiwara Y, Peng C & Orkin SH (2014) Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* **14**, 68–80.
- 74 Lee SC, Miller S, Hyland C, Kauppi M, Lebois M, Di Rago L, Metcalf D, Kinkel SA, Josefsson EC, Blewitt ME *et al.* (2015) Polycomb repressive complex 2 component Suz12 is required for hematopoietic stem cell function and lymphopoiesis. *Blood* **126**, 167–175.
- 75 Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J *et al.* (2010) DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424–2433.
- 76 Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A, Kosmider O, Le Couedic JP, Robert F, Alberdi A *et al.* (2009) Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289–2301.
- 77 Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, Stevens-Linders E, van Hoogen P, van Kessel AG, Raymakers RA *et al.* (2009) Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* **41**, 838–842.
- 78 Glezak MA & Seto E (2007) Histone deacetylases and cancer. *Oncogene* **26**, 5420–5432.
- 79 Rampal R, Alkaln A, Madzo J, Vasanthakumar A, Pronier E, Patel J, Li Y, Ahn J, Abdel-Wahab O, Shih A *et al.* (2014) DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell Rep* **9**, 1841–1855.
- 80 Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF *et al.* (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567.
- 81 Alkaln A, Garrett-Bakelman FE, Kormaksson M, Busuttill J, Zhang L, Khrebtukova I, Milne TA, Huang Y, Biswas D, Hess JL *et al.* (2012) Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet* **8**, e1002781.
- 82 Pelegri F & Lehmann R (1994) A role of polycomb group genes in the regulation of gap gene expression in Drosophila. *Genetics* **136**, 1341–1353.
- 83 Beuchle D, Struhl G & Muller J (2001) Polycomb group proteins and heritable silencing of Drosophila Hox genes. *Development* **128**, 993–1004.

- 84 Yu JR, Lee CH, Oksuz O, Stafford JM & Reinberg D (2019) PRC2 is high maintenance. *Genes Dev* **33**, 903–935.
- 85 Schuettengruber B, Bourbon HM, Di Croce L & Cavalli G (2017) Genome regulation by polycomb and trithorax: 70 years and counting. *Cell* **171**, 34–57.
- 86 Ntziachristos P, Tsirigos A, Van Vlierberghe P, Nedjic J, Trimarchi T, Flaherty MS, Ferres-Marco D, da Ros V, Tang Z, Siegle J *et al.* (2012) Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med* **18**, 298–301.
- 87 Roels J, Kuchmiy A, De Decker M, Strubbe S, Lavaert M, Liang KL, Leclercq G, Vandekerckhove B, Van Nieuwerburgh F, Van Vlierberghe P *et al.* (2020) Distinct and temporary-restricted epigenetic mechanisms regulate human alphabeta and gammadelta T cell development. *Nat Immunol* **21**, 1280–1292.
- 88 Bond J, Labis E, Marceau-Renaut A, Duployez N, Labopin M, Hypolite G, Michel G, Ducassou S, Boutroux H, Nelken B *et al.* (2018) Polycomb repressive complex 2 haploinsufficiency identifies a high-risk subgroup of pediatric acute myeloid leukemia. *Leukemia* **32**, 1878–1882.
- 89 Neumann M, Greif PA & Baldus CD (2013) Mutational landscape of adult ETP-ALL. *Oncotarget* **4**, 954–955.
- 90 Yuan XQ, Peng L, Zeng WJ, Jiang BY, Li GC & Chen XP (2016) DNMT3A R882 mutations predict a poor prognosis in AML: a meta-analysis from 4474 patients. *Medicine* **95**, e3519.
- 91 Gale RE, Lamb K, Allen C, El-Sharkawi D, Stowe C, Jenkinson S, Tinsley S, Dickson G, Burnett AK, Hills RK *et al.* (2015) Simpson's paradox and the impact of different DNMT3A mutations on outcome in younger adults with acute myeloid leukemia. *J Clin Oncol* **33**, 2072–2083.
- 92 Bond J, Touzart A, Lepretre S, Graux C, Bargetzi M, Lhermitte L, Hypolite G, Leguay T, Hicheri Y, Guillermin G *et al.* (2019) DNMT3A mutation is associated with increased age and adverse outcome in adult T-acute lymphoblastic leukemia. *Haematologica* **104**, 1617–1625.
- 93 Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M *et al.* (2014) Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**, 2477–2487.
- 94 Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burt N, Chavez A *et al.* (2014) Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488–2498.
- 95 Liu S, Cadoux-Hudson T & Schofield CJ (2020) Isocitrate dehydrogenase variants in cancer – cellular consequences and therapeutic opportunities. *Curr Opin Chem Biol* **57**, 122–134.
- 96 Takao S, Forbes L, Uni M, Cheng S, Pineda JMB, Tarumoto Y, Cifani P, Minuesa G, Chen C, Kharas MG *et al.* (2021) Convergent organization of aberrant MYB complex controls oncogenic gene expression in acute myeloid leukemia. *eLife* **10**, e65905.
- 97 Sanda T, Lawton LN, Barrasa MI, Fan ZP, Kohlhammer H, Gutierrez A, Ma W, Taterek J, Ahn Y, Kelliher MA *et al.* (2012) Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209–221.
- 98 Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, Etchin J, Lawton L, Sallan SE, Silverman LB *et al.* (2014) Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377.
- 99 Todd MA & Picketts DJ (2012) PHF6 interacts with the nucleosome remodeling and deacetylation (NuRD) complex. *J Proteome Res* **11**, 4326–4337.
- 100 Van Vlierberghe P, Palomero T, Khiabanian H, Van der Meulen J, Castillo M, Van Roy N, De Moerloose B, Philippe J, Gonzalez-Garcia S, Toribio ML *et al.* (2010) PHF6 mutations in T-cell acute lymphoblastic leukemia. *Nat Genet* **42**, 338–342.
- 101 Van Vlierberghe P, Patel J, Abdel-Wahab O, Lobry C, Hedvat CV, Balbin M, Nicolas C, Payer AR, Fernandez HF, Tallman MS *et al.* (2011) PHF6 mutations in adult acute myeloid leukemia. *Leukemia* **25**, 130–134.
- 102 Mori T, Nagata Y, Makishima H, Sanada M, Shiozawa Y, Kon A, Yoshizato T, Sato-Otsubo A, Kataoka K, Shiraishi Y *et al.* (2016) Somatic PHF6 mutations in 1760 cases with various myeloid neoplasms. *Leukemia* **30**, 2270–2273.
- 103 Soto-Feliciano YM, Bartlebaugh JME, Liu Y, Sanchez-Rivera FJ, Bhutkar A, Weintraub AS, Buenrostro JD, Cheng CS, Regev A, Jacks TE *et al.* (2017) PHF6 regulates phenotypic plasticity through chromatin organization within lineage-specific genes. *Genes Dev* **31**, 973–989.
- 104 Stanulović VS, Binhasan S, Dorrington I, Ward DG & Hoogenkamp M (2020) PHF6 interacts with LMO2 during normal haematopoiesis and in leukaemia and regulates gene expression and genome integrity. *bioRxiv* 2020.08.18.255471.
- 105 Li M, Xiao L, Xu J, Zhang R, Guo J, Olson J, Wu Y, Li J, Song C & Ge Z (2016) Co-existence of PHF6 and NOTCH1 mutations in adult T-cell acute lymphoblastic leukemia. *Oncol Lett* **12**, 16–22.
- 106 Zhang H, Wang H, Qian X, Gao S, Xia J, Liu J, Cheng Y, Man J & Zhai X (2020) Genetic mutational analysis of pediatric acute lymphoblastic leukemia

- from a single center in China using exon sequencing. *BMC Cancer* **20**, 211.
- 107 Bond J, Krzywon A, Lhermitte L, Roumier C, Roggy A, Belhocine M, Kheirallah AA, Villarese P, Hypolite G, Garnache-Ottou F *et al.* (2020) A transcriptomic continuum of differentiation arrest identifies myeloid interface acute leukemias with poor prognosis. *Leukemia* **35**, 724–736.
 - 108 Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, Beverloo HB, Moorhouse MJ, van der Spek PJ, Lowenberg B *et al.* (2004) Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* **350**, 1617–1628.
 - 109 Haferlach T, Kohlmann A, Bacher U, Schnittger S, Haferlach C & Kern W (2007) Gene expression profiling for the diagnosis of acute leukaemia. *Br J Cancer* **96**, 535–540.
 - 110 Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, Behm FG, Pui CH, Downing JR, Gilliland DG *et al.* (2002) Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75–87.
 - 111 Soulier J, Clappier E, Cayuela JM, Regnault A, Garcia-Peydro M, Dombret H, Baruchel A, Toribio ML & Sigaux F (2005) HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* **106**, 274–286.
 - 112 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550.
 - 113 Laurenti E & Gottgens B (2018) From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426.
 - 114 Jacobsen SEW & Nerlov C (2019) Haematopoiesis in the era of advanced single-cell technologies. *Nat Cell Biol* **21**, 2–8.
 - 115 van der Maaten L & Hinton G (2008) Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* **9**, 2579–2605.
 - 116 Kobak D & Berens P (2019) The art of using t-SNE for single-cell transcriptomics. *Nat Commun* **10**, 5416.
 - 117 Ferrell PB Jr, Diggins KE, Polikowsky HG, Mohan SR, Seegmiller AC & Irish JM (2016) High-dimensional analysis of acute myeloid leukemia reveals phenotypic changes in persistent cells during induction therapy. *PLoS One* **11**, e0153207.
 - 118 Mehtonen J, Polonen P, Hayrynen S, Dufva O, Lin J, Liuksiala T, Granberg K, Lohi O, Hautamaki V, Nykter M *et al.* (2019) Data-driven characterization of molecular phenotypes across heterogeneous sample collections. *Nucleic Acids Res* **47**, e76.
 - 119 Fornerod M, Zwaan CM, Zhang J, Pounds S, Downing JR, Klcio JM, Hasle H, Locatelli F, van den Heuvel-Eibrink MM, Reinhardt D *et al.* (2019) Integrative analysis of pediatric acute leukemia identifies immature subtypes that span a T lineage and myeloid continuum with distinct prognoses. *Blood* **134**, 918.
 - 120 Gu Z, Churchman ML, Roberts KG, Moore I, Zhou X, Nakitandwe J, Hagiwara K, Pelletier S, Gingras S, Berns H *et al.* (2019) PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet* **51**, 296–307.
 - 121 Geron I, Savino AM, Tal N, Brown J, Turati V, James C, Sarno J, Lee YN, Gil AR, Fishman H *et al.* (2020) An instructive role for IL7RA in the development of human B-cell precursor leukemia. *bioRxiv* 2020.01.27.919951.
 - 122 Montefiori L, Seliger S, Gu Z, Ma X, Xu B, Chen X, Dickerson K, Westover T, Stengel A, Iacobucci I *et al.* (2020) Enhancer hijacking of BCL11B defines a subtype of lineage ambiguous acute leukemia. *Blood* **136**, LBA-3-LBA-3.
 - 123 Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H & Grimes HL (2016) Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* **537**, 698–702.
 - 124 Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland MC, Heinecke A, Radmacher M, Marcucci G, Whitman SP *et al.* (2008) An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood* **112**, 4193–4201.
 - 125 Castaigne S, Pautas C, Terre C, Raffoux E, Bordessoule D, Bastie JN, Legrand O, Thomas X, Turlure P, Reman O *et al.* (2012) Effect of gemtuzumab ozogamicin on survival of adult patients with de-novo acute myeloid leukaemia (ALFA-0701): a randomised, open-label, phase 3 study. *Lancet* **379**, 1508–1516.
 - 126 Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R & Smibert P (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865–868.
 - 127 Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, Olsen BN, Mumbach MR, Pierce SE, Corces MR *et al.* (2019) Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925–936.
 - 128 Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, Parks B, Gars E, Liedtke M, Zheng GXY *et al.* (2019) Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* **37**, 1458–1465.

- 129 Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F & Newell EW (2018) Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38–44.
- 130 Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho YJ, Klein A, Hofmann O & Camargo FD (2014) Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327.
- 131 Busch K, Klapproth K, Barile M, Flossdorf M, Holland-Letz T, Schlenner SM, Reth M, Hofer T & Rodewald HR (2015) Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* **518**, 542–546.
- 132 Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, Offner S, Dunant CF, Eshkind L, Bockamp E *et al.* (2008) Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129.